Chemical Effects in Biological Systems (CEBS)

The CEBS Knowledge Base

Concepts, Standards, Challenges, Collaborations

Michael D. Waters, Ph.D.
Assistant Director for Database Development
National Center for Toxicogenomics, NIEHS, NIH, DHHS,
Research Triangle Park, NC 27709

EPA Science Forum
2 June 2004

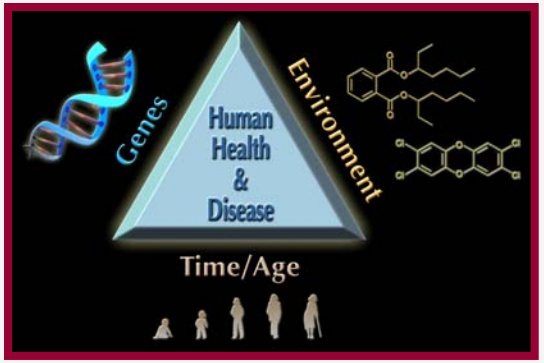
National Institute of Environmental Health Sciences
National Institutes of Health
Department of Health and Human Services

Acknowledgements

Toxicogenomics Defined

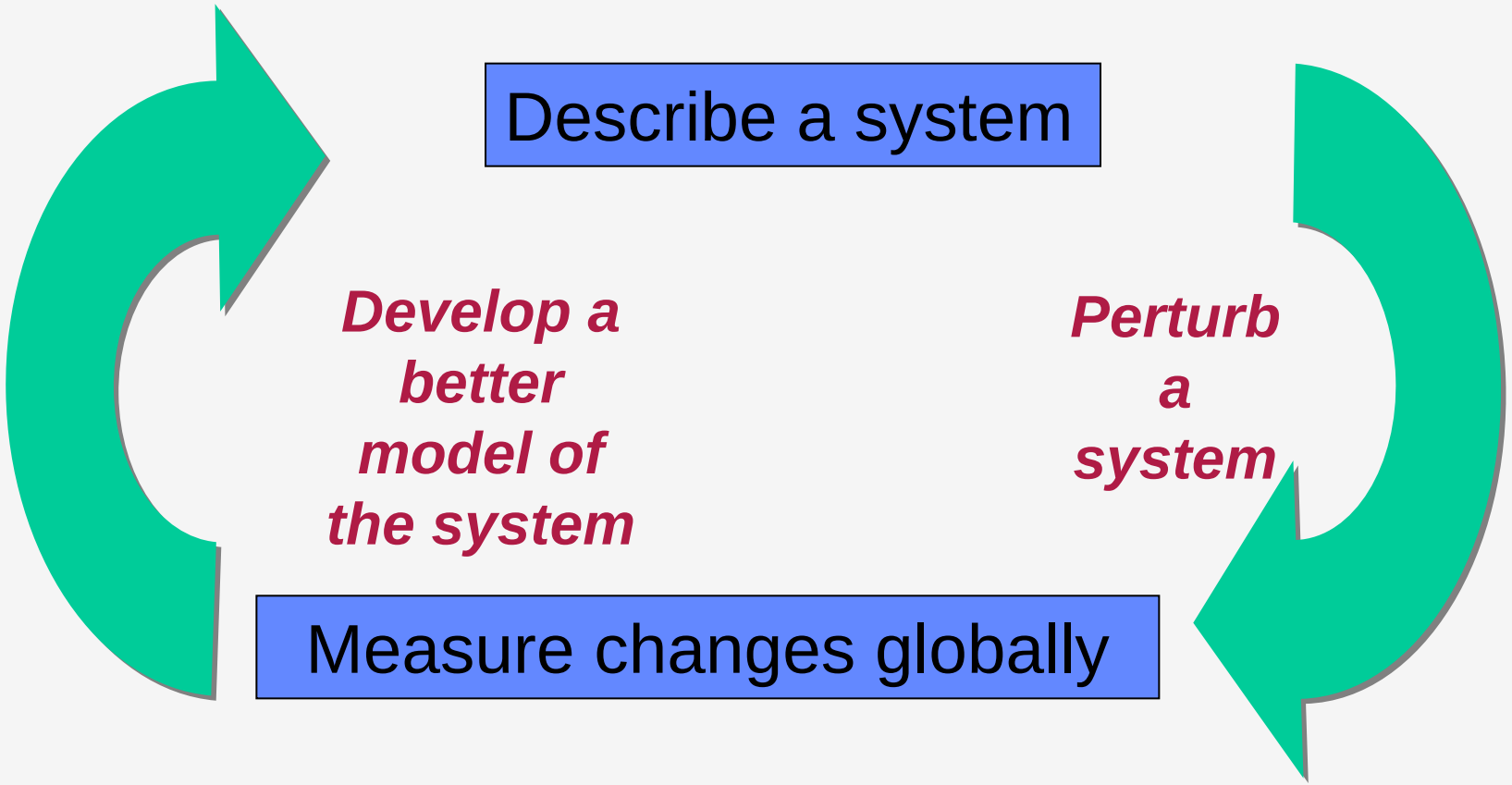
Toxicogenomics is the study of the response of a genome to environmental stressors and toxicants.

- Combines genetics, genomic-scale mRNA expression (transcriptomics), cell and tissue-wide protein expression (proteomics), metabolite profiling (metabonomics), and bioinformatics with conventional toxicology in an effort to understand the role of gene-environment interactions in disease.



Systems Toxicology: a complete description of the toxicological interactions within a system

Ideker, Galitski, & Hood (2001) A new approach to decoding life: systems biology. *Ann Rev Genomics Hum Genet* 2: 343-372.



Waters, et al. (2003) Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *Environ Health Perspectives* 111: 811-824.

A Knowledge Base Defined

*A **knowledge base** uses **data** and **information** to carry out **tasks** that create **new information and new understanding**.*

- The CEBS knowledge base aims to be a dynamic system for integrating large volumes of disparate information in a framework that serves as a continually changing heuristic engine.

- CEBS will evolve both in content and capabilities to become a “system of predictive toxicology.”

Toxicogenomics Objectives

Why do we need CEBS?

Compare toxicogenomic effects of chemicals/ stressors across species

- Yielding signatures of altered gene/protein expression

"Phenotypically anchor" these changes with conventional toxicology data

- Classifying effects as well as disease phenotypes

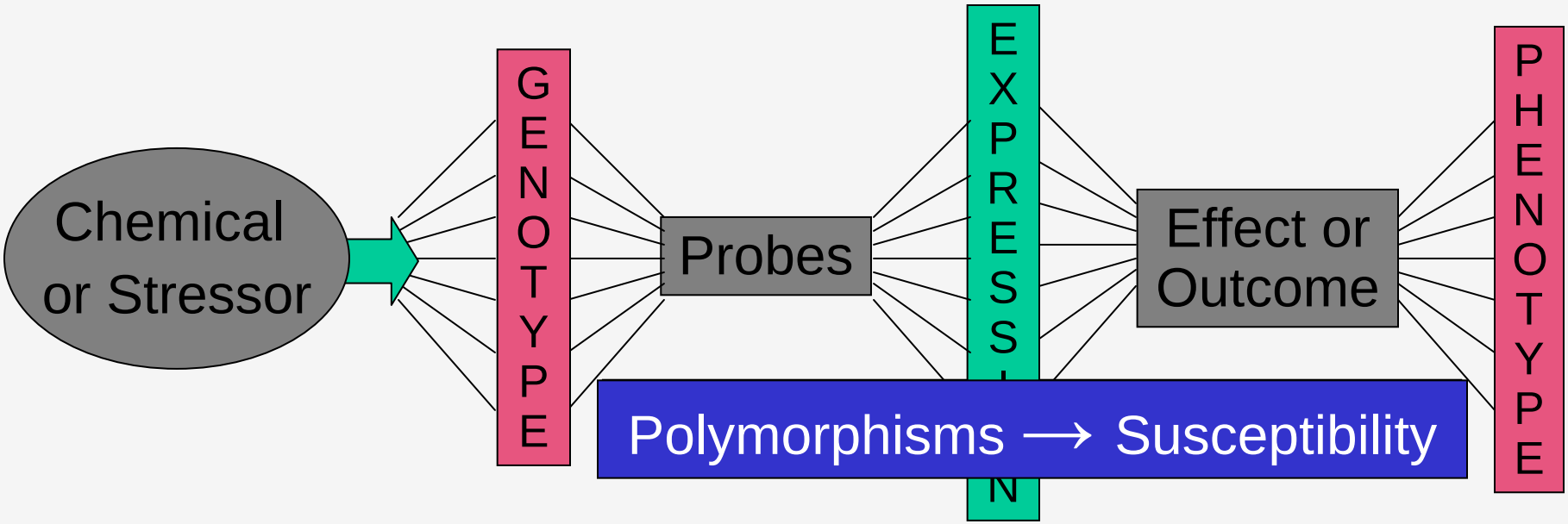
Delineate global changes as *adaptive, pharmacologic or toxic outcomes*

- Defining biomarkers, sequence of key events, modes/mechanisms of action

Two Hallmarks of CEBS

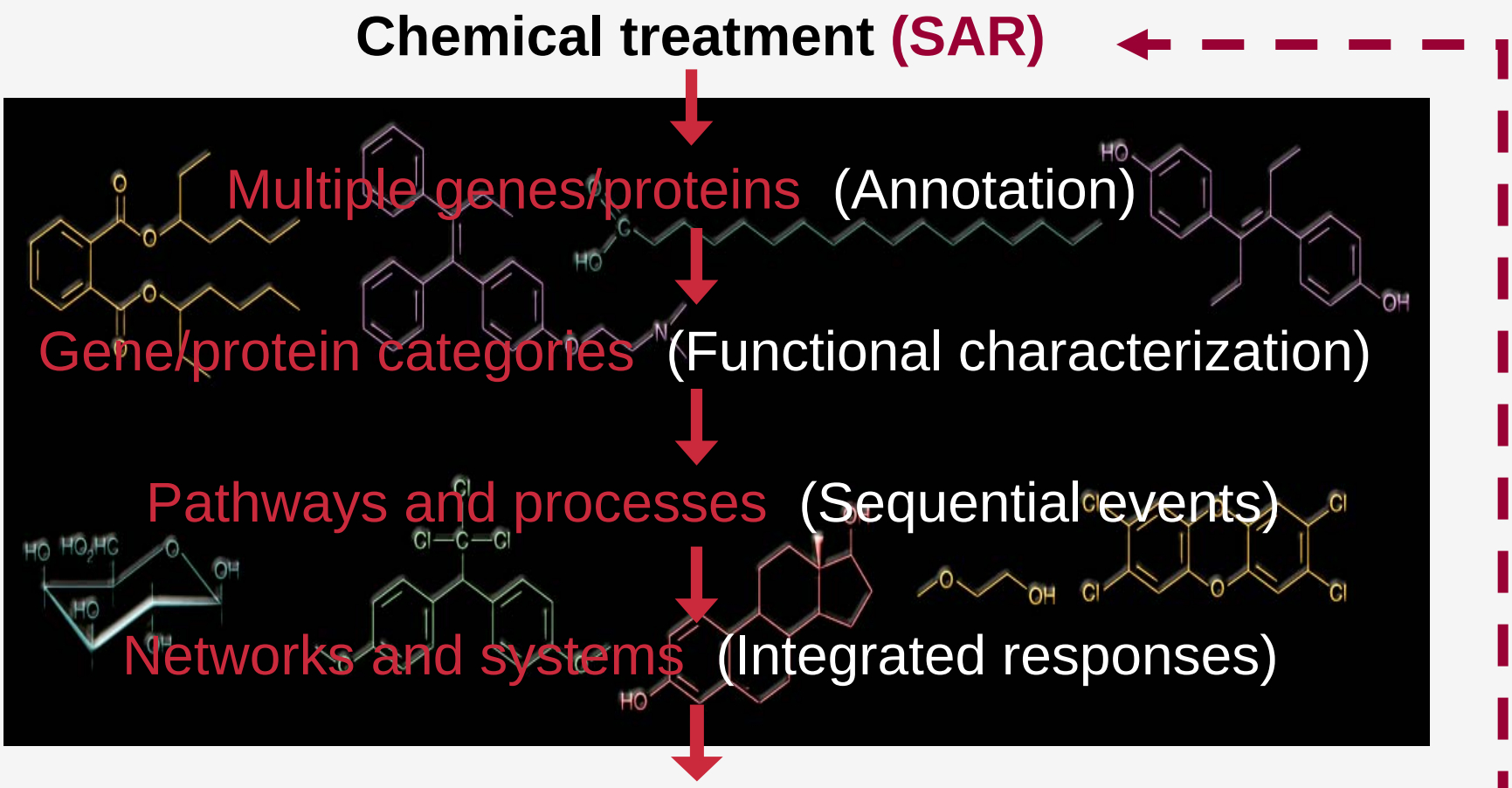
Sequence Anchoring:
Probes are anchored in genomic sequence (chromosome coordinates)

Phenotypic anchoring:
Toxicological effects (expression profiles) are anchored in phenotype using controlled vocabulary.



Bioinformatics & Interpretive Challenges

Opportunities in building a Knowledge Base



Immediate Objective

Establish a Database to:

- Capture, store and analyze gene expression data produced from toxicogenomic experiments in different laboratories
- Interrogate gene expression data using queries from genomic, experimental and toxicological domains
- Gain knowledge of relationships *between gene expression changes and toxicological endpoints*

Main Challenge

- Provide internally consistent data, and allow comparability among many toxicogenomic experiments – standards are essential

Toward Public Data Exchange

NTP

TDMS and
ClinChem DBs

NCT
National Center for Toxicogenomics

CEBS

Pharmaco-genomic Data Standards

HL7/CDISC/IC3

EMBL-EBI
ILSI
International Life Sciences INSTITUTE

Tox/ArrayExpress

meta-experiment

subject factors

simple observations

complex observations

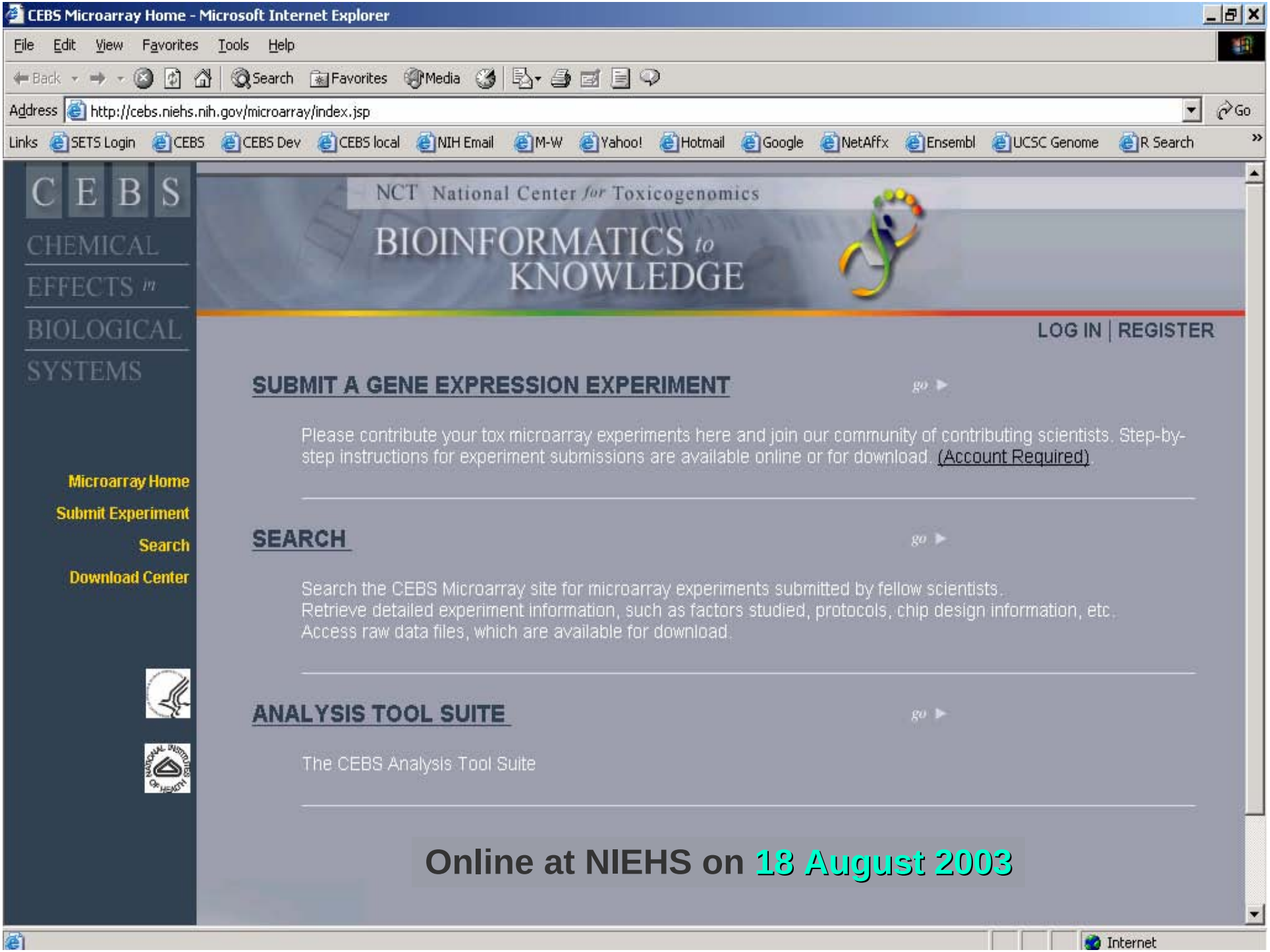
Xybion Medical Systems

Data Off-load
(flat files)

Array Track
version 2.02
NCTR

SEND will ultimately replace current FDA guidance on submitting electronic data to the Agency

SEND guidelines
version 1.5 4/14/04


CEBS Microarray Home - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://cebs.niehs.nih.gov/microarray/index.jsp
Go
Links SETS Login CEBS CEBS Dev CEBS local NIH Email M-W Yahoo! Hotmail Google NetAffx Ensembl UCSC Genome R Search
C E B S
CHEMICAL EFFECTS in BIOLOGICAL SYSTEMS
NCT National Center for Toxicogenomics
BIOINFORMATICS to KNOWLEDGE
LOG IN | REGISTER
Microarray Home
Submit Experiment
Search
Download Center
SUBMIT A GENE EXPRESSION EXPERIMENT
go
Please contribute your tox microarray experiments here and join our community of contributing scientists. Step-by-step instructions for experiment submissions are available online or for download. (Account Required)
SEARCH
go
Search the CEBS Microarray site for microarray experiments submitted by fellow scientists.
Retrieve detailed experiment information, such as factors studied, protocols, chip design information, etc.
Access raw data files, which are available for download.
ANALYSIS TOOL SUITE
go
The CEBS Analysis Tool Suite
Online at NIEHS on 18 August 2003
Internet

Compare Arrays to Identify Differentially Expressed Gene(s)

Data Processing Options

Minimum Fold Change: 2

Statistical Significance: (only applicable when comparing multiple samples/arrays)

1. Apply the following test for each gene:
   - (•) t test: Welch's two sample t test
   - ( ) Wilcoxon Test: Mann-Whitney test
2. Choose a *p*-value threshold and a multiple testing procedure to apply
   - (•) Directly use single gene test *p*-values at threshold of 0.05
   - ( ) Control False Discovery Rate (*FDR*) with adjusted *p*-value below 0.1

     Method to control *FDR*:
     - (•) Benjamini & Hochberg step-up procedure
     - ( ) Benjamini & Yekutieli step-up procedure
   - ( ) Control Family-wise Type-I Error Rate (*FWER*) with adjusted *p*-value below 0.1

     Method to control *FWER*:
     - (•) Holm step-down procedure
     - ( ) Sidak single-step procedure
     - ( ) Sidak step-down procedure
     - ( ) Bonferroni
   - ( ) Do not use the *p*-values for gene selection

Biological Analysis of Gene Expression Data

Data Viewer Options

analysis. Biological annotation and information about categories of biological activities (CBA) from different sources are incorporated during this stage of analysis to facilitate understanding of gene expression data.

- ( ) **Perform Gene Category Analysis for BioCarta Pathways**
- ( ) **Perform Gene Category Analysis for KEGG Pathways**
- (•) **View Expression Report for All Differentially Expressed Genes**

Summary of Differentially Expressed Genes

Gene Category Name	Total	Up	Down	Change	Enrichment	Fisher Exact Test p-value	View Detailed Expression Reports
Cyclins and Cell Cycle Regulation	18	3	0	3	2.47619	1.3E-4	Genes | Diagram
B Cell Survival Pathway	10	2	0	2	2.97143	0.00137	Genes
PTEN dependent Cell Cycle Arrest and Apoptosis	14	2	0	2	2.12245	0.00273	Genes | Diagram
Inhibition of Cellular Proliferation by Gleevec	17	2	0	2	1.7479	0.00403	Genes
Influence of Ras and Rho proteins on G1 to S Transition	21	2	0	2	1.41497	0.00613	Genes
Cell Cycle: G1/S Check Point	23	2	0	2	1.29193	0.00734	Genes | Diagram
IL-2 Receptor Beta Chain in T cell Activation	27	2	0	2	1.10053	0.01003	Genes
GATA3 Participate in Activating the Th2 Cytokine Genes Expression	4	1	0	1	3.71429	0.0223	Genes | Diagram
TSP-1 Induced Apoptosis in Microvascular Endothelial Cell	5	1	0	1	2.97143	0.0278	Genes | Diagram
CDK Regulation of DNA Replication	7	1	0	1	2.12245	0.0387	Genes
Phospholipase C Signaling Pathway	7	1	0	1	2.12245	0.0387	Genes | Diagram
The IGF-1 Receptor and Longevity	8	1	0	1	1.85714	0.0441	Genes

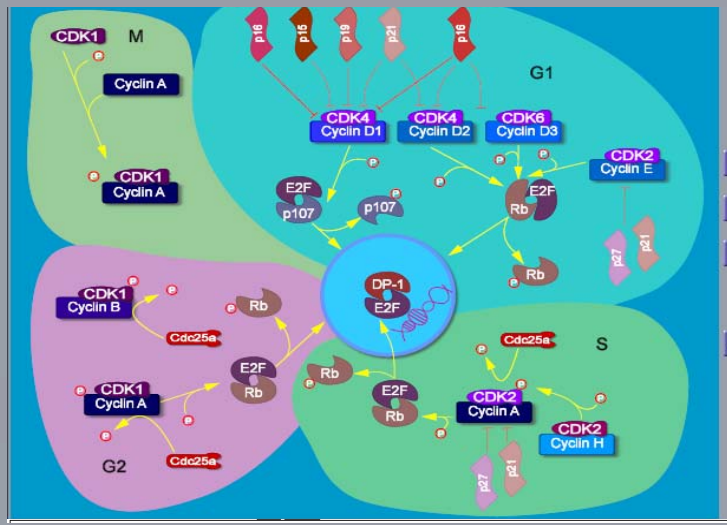
Annotation for Selected Expressed Genes

Gene Information For: Mm Adamts1 a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 1

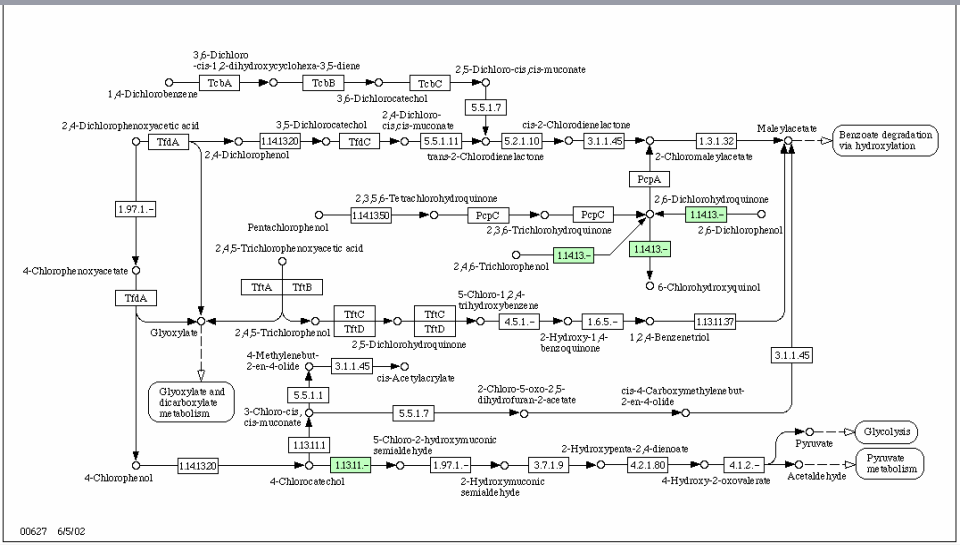
Sequence ID: BB042202, NM_009621, BB135256

Database Links

Visualization on BioCarta Pathways

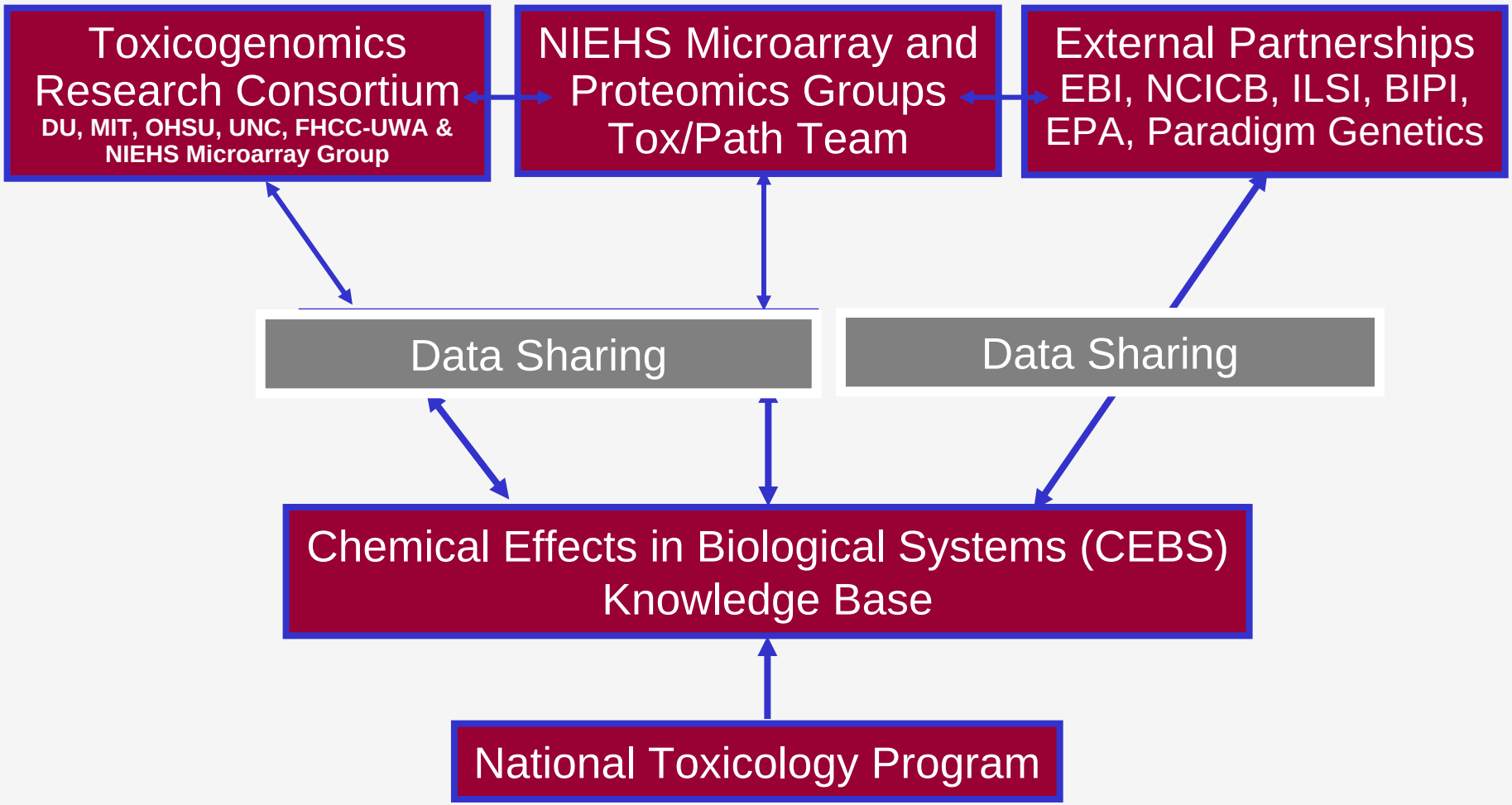


Visualization on KEGG Pathways



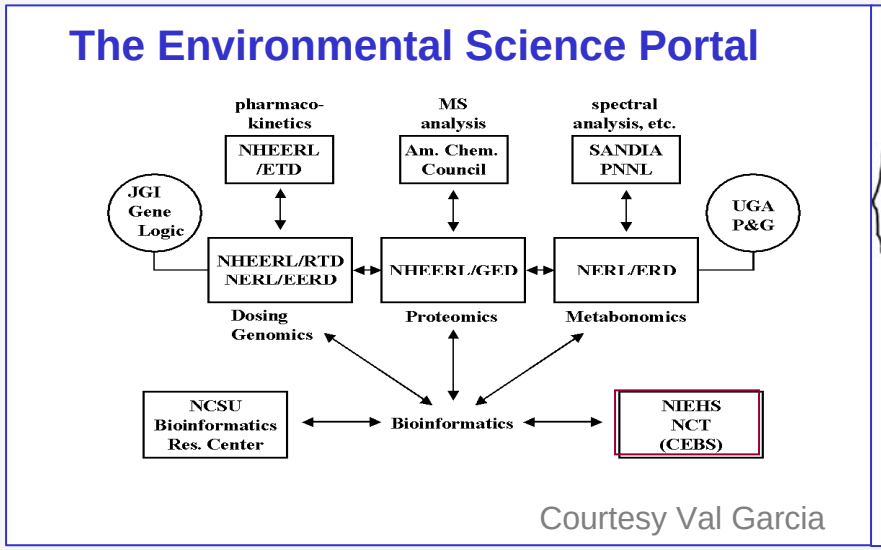
Now Gaining Content for CEBS

Intramural and Extramural Partnerships

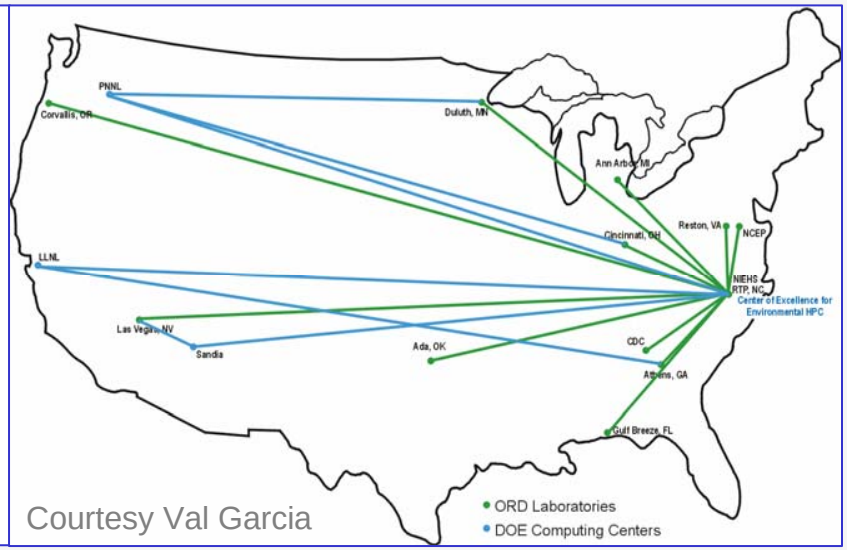


NCT- EPA Collaborations (CEBS)

- Metabonomics Center of Excellence with NERL.
- SAR Interface with DSSTox, Research in Toxicogenomics, Computational Toxicology with NHEERL.
- Toxicogenomics Applications in Risk Assessment with NCEA.



Courtesy Val Garcia



Courtesy Val Garcia

RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

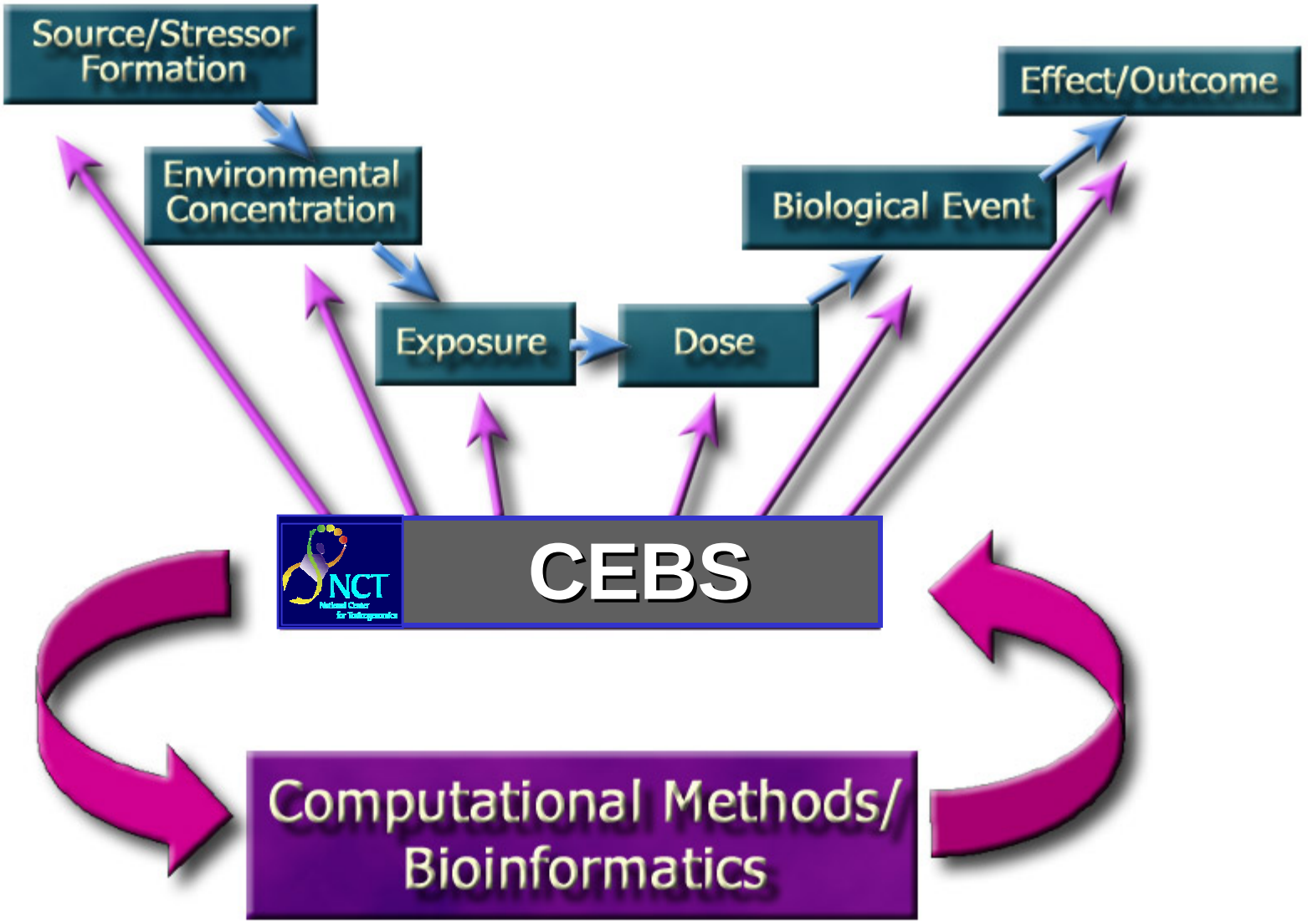
To integrate modern computing and information technology with the technology of molecular biology and chemistry to improve EPA's prioritization of data requirements and risk assessments for toxic chemicals

COMPUTATIONAL TOXICOLOGY

Courtesy Robert Kavlock

RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions



COMPUTATIONAL TOXICOLOGY

Courtesy Robert Kavlock

Conclusions

- Toxicogenomics will change the way toxicology is performed.
- Toxicogenomics will contribute new methods, new data, and new interpretation to environmental toxicology.
- CEBS will be a key component in toxicological interpretation – linking transcriptomics, proteomics, metabonomics, and toxicology to generate new knowledge.

Publications 2003-04

Waters, MD, Boorman G, Bushel P, Cunningham M, Irwin R, Merrick A, Olden K, Paules R, Selkirk J, Stasiewicz S, Weis B, Van Houten B, Walker N, and Tennant R, Systems toxicology and the chemical effects in biological systems knowledge base, *Environmental Health Perspectives* 111, 811-824 (2003).

Waters, MD, J.K. Selkirk, and K. Olden, The Impact of new technologies on human population studies, *Mutation Research* 544, 349-360 (2003).

Waters, MD, K. Olden, and R.W. Tennant, Toxicogenomic approach for assessing toxicant-related disease, *Mutation Research* 544, 415-424 (2003).

Mattes WB, Pettit SD, Sansone A, Bushel PR, and Waters MD, Database development in toxicogenomics: issues and efforts, *Environmental Health Perspectives* 112, 495-505 (2004).

Xirasagar S, Gustafson S, Merrick AB, Tomer KB, Stasiewicz S, Chan DD, Yost KJ, Yates JR, Xiao N, Waters MD, CEBS object model for systems biology data, CEBS SysBio-OM, *Bioinformatics,* in press (2004).